Applied Network Science

Quantitative analysis of trade networks: data and robustness

Najmeh Sajedianfard[1], Ebrahim Hadian[1*], Ali Hussain Samadi[1], Zahra Dehghan Shabani[1], Somwrita Sarkar[2] and P. A. Robinson[3,4]

*Correspondence:
ehadian@rose.shirazu.ac.ir
[1] Department of Economics,
Shiraz University, Shiraz, Iran
Full list of author information
is available at the end of the
article

Abstract

A common issue in trade network analysis is missing data, as some countries do not report trade flows. This paper explores what constitutes suitable data, how to deal with missing data, and demonstrates the results using key network measures. All-to-all potential connectivity of trade between countries is considered as a starting point, in contrast to the common approach of analyzing trade networks using only the countries that actually report trade flows. In order to fill the gap between the two approaches, a more complete dataset than just the dataset of trade between reporting countries is reconstructed and the robustness of studying this bigger dataset is examined. The difference between imputed and actual network adjacency matrices is evaluated based on several centrality measures. The results are illustrated using ten commodity groups from the United Nations Database, which demonstrate that under the proposed reconstruction procedure the ranks of the countries do not change significantly as the size of the imputed network becomes bigger or smaller. Further, the degree distributions of networks based on reporting countries and trading partners are the same to within their uncertainties. So, it is robust to study the imputed bigger network that provides richer insights into trade relations, particularly for nonreporting countries.

Keywords: Centrality, Data analysis, Entropy, Principal component analysis, Trade networks

Introduction

Is it robust to study the bilateral trade between countries in a network consisting of as many countries and territories as possible in the world trade web? Studying a country's situation in the world trade network is the primary stage of future policy making for that given country. Therefore, regardless of countries' sizes, this matters for all sets of under developed, developing, and developed countries.

Network theory has been widely used in studying economic interactions [1–3].The intricacies of trading relations and their inter-country distributions within the world trade network has been successfully captured and analyzed by applying network concepts. In the last two decades several studies have been carried out on the topology and structure of trade networks [4–17], and their evolution and dynamical behavior [18–27].

Few researchers have addressed the structure of trade networks by taking into account as many countries as possible involved in trade in the whole world; most focused on the biggest economies and have often neglected the other countries. However, these countries position in trade networks is important to their citizens (often in the billions in aggregate) and neighbors. This is while it is important to study a trade network that considers *as many bilateral trade as possible* that happens in trade networks. In other words, it is necessary to study an expanded data and not just rely on the available data [28]. In order to fulfill all countries' goal of understanding their status in trade networks, it is necessary to consider as many countries in the trade network as possible rather than just a subset.

In order to address the above issue, one must use real data. The main source for trade statistics is the United Nations Commodity Trade Statistics Database. However, a crucial issue is that there are missing data because in each given year, a number of countries have not reported their trade flows due to delays in reporting, disruptions (war, disaster) in the country, or being a too small or sparsely populated to support reporting infrastructure. Thus, comparing trade networks longitudinally is problematic since the size of the network is not always the same in different years. Nonetheless, accounting for trade that involves nonreporting countries is important—not least to those countries themselves. Because, roughly half of the world's population belongs to the nonreporting countries [29]. Therefore, constructing same-sized networks and robustly imputing missing data are issues that need addressing.

After collecting the data for each country in each year, it is found out that each country reports its trade flows with a list of trading partners at the world level, at the level of some special areas, regions, and categories, and at the level of individual countries. Therefore, one way to impute missing data and construct same-sized networks is to use data from the *trading partners* of the reporting countries, some of which may report trade with a country that did not itself report. Following [28, 30], in this paper, exports of a country is used as a proxy for the import of another country from the first country in the import network and conversely in the export network.

In this paper, we extend results within the framework of network theory. So, the main focus of this paper is on the missing data due to lack of reporting trade flows, proposing network link imputation and reconstruction, and testing the robustness of network analysis on the imputed and reconstructed bigger network containing as many trading partners as can be determined. The contribution of this study is that, the validity of studying an expanded data set based on network measures is verified, which was, to our knowledge, not verified in previous studies. In other words, the main contributions are that a systematic way to impute missing link data is presented, a network is reconstructed, and then by using centrality measures as key test cases, the robustness of using the network with imputed data for trade network analysis is tested.

In order to fill the gaps of consistent network size for longitudinal studies and imputing missing data, this paper analyzes the size of the network to be studied based on the trading partners of each reporting country. Also, some centrality measures, degree, and degree distribution are used to quantify the properties of networks. This analysis is done by using the trade flow (import and export) data of ten commodity groups [based on

the Standard International Trade Classification (SITC)] in 2016 from the United Nations Commodity Trade (UN Comtrade) Statistics database.

The rest of the paper is organized as follows. "Materials and methods" section contains the network concepts, the database, and the methodology used in this paper. "Results and discussion" section presents the results and discussion. Finally, the conclusions are discussed in the last section.

Materials and methods
Network concepts
In this section relevant network concepts are represented. First, the adjacency matrix is introduced, a central concept in network theory. Second, the degree and degree distribution are explained. Third, the centrality measures are studied and the corresponding economic interpretation is given.

Adjacency matrix
One of the fundamental mathematical structures in network theory is the adjacency matrix, which represents the connections between nodes in a network. Equation (1) shows an adjacency matrix for N countries:

$$A^{(N)} = \begin{pmatrix} a_{11} & \cdots & a_{1P} & a_{1,P+1} & \cdots & a_{1N} \\ a_{21} & \cdots & a_{2P} & a_{2,P+1} & \cdots & a_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{P1} & \cdots & a_{PP} & a_{P,P+1} & \cdots & a_{PN} \\ a_{P+1,1} & \cdots & a_{P+1,P} & a_{P+1,P+1} & \cdots & a_{P+1,N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{N1} & \cdots & a_{NP} & a_{N,P+1} & \cdots & a_{NN} \end{pmatrix}. \tag{1}$$

In this paper, the trade flows between countries are considered. Because trade flows between countries are not symmetric (i.e., import (export) of country i from (to) country j is not necessarily equal to import (export) of country j from (to) country i), and countries do not import from themselves, the network in this study is a weighted directed network without self-loops. Therefore, the adjacency matrix of this network is an asymmetric adjacency matrix with zero diagonal elements. In the import network, a_{ij} represents the import value of country j from country i and in the export network, the a_{ij} represents export of country i to country j.

In this paper, the 2016 trade flow data is studied. In 2016, only 154 countries provided their reports. After studying the trading partners of those countries, 234 trading units, including the 154 countries, were found to be the trading partners. So, a 234-by-234 adjacency matrix is reconstructed with 154 reporting countries labeling the first 154 columns without the loss of generality and the 234 trading partner countries labeling the rows.

Degree (local centrality measure) and degree distribution
Degree is the total number of links incident on a node, or the number of links a node is connected to. In weighted networks, it is the sum of the weights of the links which is

called *strength*. In directed networks, the concept of the degree is subdivided to *in-* and *out-*degree which are defined as

$$d_i^{in} = \sum_{j=1}^{N} a_{ji},$$

(2)

$$d_i^{out} = \sum_{j=1}^{N} a_{ij},$$

(3)

where d_i^{in} stands for in-degree of node i, d_i^{out} stands for out-degree of node i, and a_{ji} and a_{ij} are elements of the corresponding adjacency matrix with $a_{jj} = 0$ for all j [3, 31].

Degree distribution: The fraction of nodes that have degree k is given by the degree distribution, which is a characteristic of the network structure and can have different functional forms [2, 32–34]. For example, random networks have binomial degree distribution which is approximately a Poisson distribution for large networks. In regular networks where all nodes have the same degree, their distribution is a delta function at that degree. Random networks and regular networks are two extreme graphs in network theory and small-world networks have properties between these two extremes [35]. Other types of networks that have a specific degree distribution include scale-free networks which have a power-law degree distribution [36–38], with probability proportional to a negative power of the degree.

Centrality measures

A central node can be described by either how connected that node is or how influential its neighbors are [3]. For instance, based on degree centrality, a central node is a node that has many connections. Also, in-degree indicates that a central node has many inflows and out-degree assigns central node status to a node that has many outflows.

In economic terms, for a weighted directed import network of a commodity, in-degree centrality indicates how dependent a country is on other countries from which it imports that specific commodity group. Particularly, a country with a high in-degree centrality measure is a highly dependent country on importing a specific commodity group and does not satisfy its own need for that commodity group. Conversely, the out-degree centrality states the extent to which a country is a source (although not necessarily a producer) of a specific commodity group for a group of other countries. Also, for a weighted directed export network of a commodity, out-degree centrality indicates the strength of a country to be a potential supplying market of that commodity. On the other hand, in-degree centrality in an export network represents the dependence of a country on the other supplying markets of a specific commodity. Here, for simplicity, in-degree and out-degree centrality measures are considered for the import and export networks, respectively.

The centrality of a node can also be determined by its neighbors' characteristics. Therefore, a node that is central based on one type of centrality measure is not necessarily central under other such measures [3, 39]. In this subsection, the centrality measures of PageRank, hubness, and authority are introduced, which are widely used global

Sajedianfard *et al. Appl Netw Sci*   (2021) 6:46

Page 5 of 26

centrality measures. Note that *eigenvector* centrality is one of the other measures that captures the centrality of a node based on its neighbors' characteristics. However, the eigenvector centrality cannot be measured in directed networks [31], so the present study does not examine this quantity.

PageRank: In this measure the centrality of a node derives from its network neighbors' and is proportional to their centrality divided by their *out-degree*. In this case, nodes that point to many others pass only a small amount of centrality on to each of those others, even if their own centrality is high [31]. In mathematical terms the PageRank centrality, x_i, of node i is defined by

$$x_i = \alpha \sum_j a_{ij} \frac{x_j}{d_j^{out}} + \beta \tag{4}$$

where α and β are positive constants and x_i and x_j are the centralities of nodes i and j, respectively. This centrality measure is defined by two endogenous and exogenous components. The endogenous component, $\alpha \sum_j a_{ij} \cdot x_j / d_j^{out}$, is based on the network topology, and the exogenous component, β, is independent of the network structure [40].

The economic significance of this centrality measure is that a country with a high PageRank centrality is connected to a central country (i.e., a country with high PageRank centrality). Also, central countries themselves have high PageRank centrality measures indicating they are core sources in that specific commodity.

Hub and authority: Page Rank centrality assigns a country high centrality if other trading units that have trade flows to this country, have high centrality. However, in the case of *directed* networks it is possible to accord a country high centrality if it *points to* (has flows to) other trading units with high centrality. Thus, there are really two types of central countries in such a network: *authorities* are countries that are important and central points of a specific commodity group in the import sector, and *hubs* are trading units that trade with important and central points of a specific commodity and export to other important and central countries (authorities) in the network. In other words, hub countries point to authorities. Besides, an authority may also be a hub and vice versa [31]. In this study, following [13], and for simplicity, hubness is considered for the export networks and authority is used for the import networks.

Note that two other types of centrality measures, *betweenness*, and *closeness*, exist, which determine the centrality of a node based on being on the path of other nodes [3, 31]. The issue with calculating these centrality measures is that the weights of the links must be the cost of reaching a node from the other [41, 42]. Since we could not find a suitable data for the cost of trade among countries, these centrality measures are not considered in this study.

Database and methodology

The trade flow data of ten commodity groups in 2016 published by the UN Comtrade database is used in this paper [43]. The ten commodity groups are as follows: food and live animals (group 0), beverages and tobacco (group 1), crude materials, inedible, except fuels (group 2), mineral fuels, lubricants and related materials (group 3), animal and vegetable oils and fats (group 4), chemicals (group 5), manufactured goods classified chiefly by material (group 6), machinery and transport equipment (group 7), miscellaneous

Sajedianfard *et al. Appl Netw Sci*   (2021) 6:46

Page 6 of 26

manufactured articles (group 8) and commodities and transactions not classified according to kind (group 9).

The UN Comtrade database has some limitations. One of these limitations is that imports reported by one country do not coincide with exports reported by its trading partners. These differences are due to various factors; for instance, the valuation of export data is based on free on board (FOB) pricing and the valuation of import data is based on cost, insurance, freight (CIF) pricing which are international commercial terms used for sea and inland waterway transports. Some other differences arise because of including or excluding particular commodities or timing (departure of exports must always precede their arrival as imports, sometimes by weeks for seaborne trade) [44].

Based on the above-mentioned limitations, we argue that one should study the trade network for a specific commodity group at a time in import and export sectors separately. Therefore, in the present work, the ten commodity groups in 2016 in the import and export sectors as the case studies for investigating the trade network. In 2016 a total of 234 trading units (countries and territories, simply termed as *countries* for brevity) were listed as trading partners of the 154 reporting countries. This implies that country j may have reported its import from country i (element a_{ij} of the adjacency matrix), but country i may not have reported its trade with any country. Thus, if a 234-by-234 matrix is modeled, there are missing entries for countries that did not report their trade flows. This smaller 234-by-154 sized data set is used to verify the robustness of studying the trade network based on trading partners, in part by examining the largest complete subset (154-by-154) and the effects of artificially deleting data from that subset. Accordingly, in order to deal with missing data, exports of country j to country i might be used as a proxy for the import of country i from country j and conversely in the export network [28, 30]. Although this proxy is not precise, it provides an approximation to the missing data.

To address the above points, a 234-by-234 weighted directed adjacency matrix was constructed, where the a_{ij} element in the import network represents the import value for country j with country i and derived from country i's exports where absent. Also, in the export network, the a_{ij} element shows the export value from country i to country j, derived from country j's imports where the export data are absent. This approach is not exact but provides an approximate way to impute the missing entries [28, 30].

Based on the International Merchandise Trade Statistics (IMTS), Concepts and Definitions (2010), the attribution of the *partner country* for imports is the country of origin and for exports is the country of last known destination [45]. So, a row of the matrix corresponds to a country of origin (i.e., a trading partner) and a column belongs to a destination country, which is a reporting country; here, a_{ij} is the import by j from i in the import network and it is the export of i to j in the export network.

Centrality measures are a tool to study the structure of the network. In order to study the robustness of analyzing the trade network with missing data inferred from trading partners, two aspects of trade networks are examined. First, we argue that if the rank of the countries based on different degree and centrality measures does not significantly change, the structure of the network is little affected by the missing data. Second, by using principal component analysis and entropy concepts, the core countries in trade networks (import/export) of the reporting countries and trading partner countries are extracted by considering the

structural features of networks. In this study, the *core countries* are those which are the main contributors to the structure of the trade networks. We argue that if the core of the reporting countries' and trading partners' networks are the same, the dominant structure of the networks would be the same. This part of the analysis is well introduced in the last subsection. Also, the former issue is studied in two cases with a structure similar to the matrices in Eq. (5).

$$
\begin{pmatrix}
a_{11} & \cdots & a_{1P} & a_{1,P+1} & \cdots & a_{1N} \\
a_{21} & \cdots & a_{2P} & a_{2,P+1} & \cdots & a_{2N} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
a_{P1} & \cdots & a_{PP} & a_{P,P+1} & \cdots & a_{PN} \\
a_{P+1,1} & \cdots & a_{P+1,P} & a_{P+1,P+1} & \cdots & a_{P+1,N} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
a_{N1} & \cdots & a_{NP} & a_{N,P+1} & \cdots & a_{NN}
\end{pmatrix}
$$
$$A \quad (Full \quad Matrix)$$

$$\downarrow$$

$$
\begin{pmatrix}
a_{11} & \cdots & a_{1P} & 0 & \cdots & 0 \\
a_{21} & \cdots & a_{2P} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
a_{P1} & \cdots & a_{PP} & 0 & \cdots & 0 \\
a_{P+1,1} & \cdots & a_{P+1,P} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
a_{N1} & \cdots & a_{NP} & 0 & \cdots & 0
\end{pmatrix}
$$
$$B \quad (Matrix \quad with \quad Missing \quad Data) \tag{5}$$

$$\downarrow$$

$$
\begin{pmatrix}
a_{11} & \cdots & a_{1P} & 0 & \cdots & 0 \\
a_{21} & \cdots & a_{2P} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
a_{P1} & \cdots & a_{PP} & 0 & \cdots & 0 \\
0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0
\end{pmatrix}
$$
$$C \quad (Full \quad Nonzero\text{-}Reduced \quad Matrix)$$

$$\downarrow$$

$$
\begin{pmatrix}
a_{11} & \cdots & a_{1P} \\
a_{21} & \cdots & a_{2P} \\
\vdots & \vdots & \vdots \\
a_{P1} & \cdots & a_{PP}
\end{pmatrix}
$$
$$D \quad (Full \quad Submatrix)$$

Here matrix A represents a full N^2 matrix with no missing data. Matrix B represents a N^2 matrix which contains available data from P reporting countries and their N trading partners, and missing data (the columns of zeros); the leading $N \times P$ part of this matrix contains the available data and the rest shows the missing data. Matrix C highlights the leading P^2 submatrix of the N^2 matrix as the largest square matrix with no missing

Sajedianfard *et al. Appl Netw Sci*   (2021) 6:46

Page 8 of 26

data; all other entries are set to zero. Finally, matrix D is the P^2 submatrix extracted from matrix C. In all of these matrices, a_{ij} is the import by country j from country i in import networks and the export of country i to country j in export networks, which can be zero in some cases. Note that zero entries in adjacency matrices are of two types; one group represents true zeros in the reporting countries' reports, and the other arises from nonreporting. This study is mainly interested in actual cases of nonreporting, which are termed missing entries here. Since the missing entries are dominant, the difference between zeros is disregarded in this research.

Our aim is to examine the effect of missing entries in the 234^2 adjacency matrix based on the trading partners' network. Therefore, in order to study the effect of missing flows in a network based on the trading partners, two cases are considered. The first case traces the change in structural status of the 154 reporting countries in the trade network based on the 234 trading partners. The second case simulates a similar situation to the first case by considering the full trade network based on the 154 reporting countries and constructing an incomplete trade network out of it. The incomplete trade network has the same fraction of missing entries as in the trading partners' network, i.e. in the new constructed network, 52 entries are missing. As the first case, the change in structural status of the new set of countries in the trade network based on the 154 reporting countries are studied in the second case.

In these cases we examine the effect of missing data by comparing countries' rank based on degree and centrality measures. For each case, the countries in the smaller adjacency matrix [i.e., matrix D in Eq. (5)] are considered as the basis for comparing their rank among the other set of networks. In the import networks, in-degree, PageRank, and authority centrality measures are examined and in the export networks, out-degree, PageRank, and hub centrality measures are considered for the analysis. This comparison is done for the two cases by using Spearman's correlation coefficient [46]. Also, the similarity of the coefficients is tested by using the Fisher Z transformation [47, 48]. In various steps, based on the matrices in Eq. (5), these cases are explained as follows.

First case: Changes in structural status of the reporting countries

In this case, there is lack of data on bilateral trade statistics between some of the countries and territories in the list of the trading partners' countries. Therefore, based on Eq. (5), the full matrix A with $N = 234$ is not available because of the lack of availability of data and this case goes from matrix B to D.

Step 1: Available matrices in Eq. (5) in the first case In 2016, 154 countries reported their trade flows but 234 countries are reported on. So, a 234^2 matrix is available that contains missing data. Therefore, there would be 154 columns (reporting countries) and 234 rows (trading partners for those reporting countries) with no missing data. This adjacency matrix is similar to matrix B in Eq. (5) with $N = 234$ and $P = 154$. In order to study the trade relations between the reporting countries, it is necessary to eliminate the rest of the countries from the 234-country network. Therefore, the trade flows of the reporting countries with the rest of the world is set to zero; i.e., all entries in rows $P + 1$ to N are set to zero to yield a matrix like C in Eq. (5) with $N = 234$ and $P = 154$.

Thereafter, a full adjacency matrix like matrix D in Eq. (5) with $P = 154$ is constructed based solely on trade between reporting countries.

Step 2: Checking the structural position of the countries In this step, centralities of the 154 reporting countries in the trade networks based solely on the trade between those countries are measured. Accordingly, these countries structural position has been measured through their centralities. Then, these countries are ranked based on the above mentioned centrality measures in import and export networks of various commodity groups. Besides, these countries are ranked in the 234-country networks- in different sectors and commodity groups- based on the centrality measures. The next step is to compare the rank of the 154 reporting countries in the 154-country and the 234-country networks, which is done by using Spearman's correlation coefficient.

Second case: Simulation of the first case and the examination of the changes in structural status of a smaller set of countries

Our aim here is to test the effect of inferring missing data on nonreporting countries from data provided by those countries that did report. Since we do not have access to the full 234-country data, we carry out this test for trade between the 154 countries that did report, which gives the largest available full matrix. We artificially render this matrix incomplete by setting to zero the same fraction of columns as are missing from the 234-country case due to nonreporting. We then use the remaining data to estimate as many as possible trade flows for the nonreporting countries and compare the resulting network structure with that of the full 154-country network. Note that for constructing the incomplete 154-country network, first, the 154 countries are ranked descendly based on their total trade (import/export) value. Second, the trade values of top 102 countries ($\frac{154}{234} \simeq \frac{102}{154} \simeq 0.66$) are kept and the trade value of the rest of the countries are set to zero. Therefore, the incomplete 154-country matrix with 102 nonzero columns is constructed. The logic behind keeping the top 102 countries is that the major trading countries always report their trade flows. Accordingly, without the loss of generality, the incomplete 154-country matrix is constructed this way.

Step 1: Available matrices in Eq. (5) in the second case The analysis is started with the largest available full matrix like matrix A in Eq. (5) in which $N = 154$. Then, the same proportion of columns of data as is missed from the original 234^2 matrix ($\frac{154}{234} \simeq \frac{102}{154}$) is intentionally deleted in the full matrix. This yields a matrix like B in Eq. (5) with $N = 154$ and $P = 102$. To see the trade relation between countries with no missing data, it is needed to eliminate the rest of the countries from the 154-country network. Therefore, the trade flows of the 102 representative countries with the rest of the 154 countries is set to zero. Hence, a matrix like matrix C with $P = 102$ is constructed and then the square part of this matrix is extracted and a matrix like D is constructed with $P = 102$.

Step 2: Checking the structural position of the countries In this case, studying the structural position of the countries is twofold. First, the rank of 102 countries in matrix D and B are compared and secondly, the rank of those countries are compared

in matrix B and A. Simply put, centrality measures of 102 countries are calculated in import and export networks of various commodity groups; i.e., trade networks based solely on the bilateral trade between those 102 countries. Then, these countries are ranked rest on their centrality measures in import and export networks. This process is done anew in the trade network formed on the trade between reporting countries which is intentionally made incomplete; i.e., a matrix like B with $N = 154$ and $P = 102$ in Eq. (5). Further, rank of the 102 countries are compared in these two networks based on their centrality measures which contributes to a correlation coefficient like r_1.

Additionally, the centrality measure of these countries are calculated in the complete trade networks; i.e., networks build exclusively on the bilateral trade between reporting countries. The adjacency matrix of these networks are similar to matrix A in Eq. (5) with $N = 154$. Next, these countries are ranked rest on the centrality measures and the ranks are compared to those in the intentionally incomplete networks, i.e. the trade networks with an adjacency matrix like B in Eq. (5). The rank comparison contributes to a correlation coefficient like r_2. Lastly, the probability of similarity between r_1 and r_2 are tested in various commodity groups and import and export networks by using the Fisher Z transformation. That is to say, the similarity of the structural position of these countries is checked.

Comparison of the core countries in reporting countries' and trading partners' trade networks

Centrality measures capture the structural network aspects. The structure of the trade networks in the previous subsection is examined using the centrality measures. The issue with this approach is that some nodes are *intentionally* deleted from the networks. In order to study the structure of the networks with no intentional interference, a composite index out of the centrality measures is constructed and the core countries of the networks are extracted. Recall that the core countries in this study are those which are the main contributors to the structure of the trade networks.

We argue that if the core countries in trading different commodity groups in reporting countries' network is the same as the core countries in trading partners' network, the structure of the trade network is not significantly affected by the missing data. The logic behind this is that the core countries are extracted by using a composite index of the network structural measures. The composite index out of the centrality measures are constructed using principal component analysis (PCA) method. In addition, entropy is applied on the composite indices to find the core countries.

Principal Component Analysis: PCA is the oldest and the most well-known multivariate statistical technique being used in dimension reduction in all science disciplines [49–51]. One of its applications is the construction of mixed indices considering different features and variables [52].

PCA computes a linear combination of the original variables equal in number to the original ones. Suppose there is a $X_{N \times K}$ matrix with K variables (indicators) and N observations. Define $R_{K \times K}$ the correlation matrix between the K variables with λ_i $(i = 1, \ldots, K)$ as the ith eigenvalue, and $v_{K \times 1}^i$ $(i = 1, \ldots, K)$ as the ith eigenvector of the $R_{K \times K}$ correlation matrix. Assume, $\lambda_1 > \lambda_2 > \ldots > \lambda_K$, then, the principal components are derived as Eq. (6).

$$PC_i = Xv^i, i = 1, \ldots, K \tag{6}$$

where PC_i stands for the ith principal component, X is the defined matrix with K variables and N observations, and v^i is the eigenvector corresponding to the ith eigenvalue being sorted in descending order [49, 53].

Note that λ_i is also defined as the variance of the ith principal component, i.e., $\lambda_i = var(PC_i)$. consequently, the first principal component is the linear combination of the initial indicators that has the biggest variance. The second principal component is the linear combination of the initial indicators that has the second biggest variance. Similarly, the Kth principal component is the linear combination of the initial indicators that has the smallest variance. The principal components are used as the basis of a single mixed variable construction out of the original variables [50, 53].

In this study, the in-degree, PageRank, and authority centrality measures are used to capture structural network characteristics of countries in the import sector. Also, the out-degree, PageRank, and hubness are considered to extract the structural network aspects of countries in the export sector. According to the notation, there are $K = 3$ variables for each network, and $N = 154$ and $N = 234$ observations in the reporting countries' and trading partners' networks, respectively. In an attempt to compare the core countries in the reporting countries' and trading partners' import and export networks, four composite indices are constructed applying PCA to the centrality measures.

As stated above, in PCA, the lowest order components describe the most variations in the original variables. Accordingly, a new index can be constructed using the lowest order components. Following Chen and Woo (2010), the new variable is calculated as Eq. (7).

$$CI_n = \sum_{j=1}^{N} \omega_j x_j, \quad \text{and} \quad \omega_j = \frac{\sum_{i=1}^{K} \lambda_i v_j^i}{\sum_{i=1}^{K} \lambda_i} \tag{7}$$

where CI_n is the composite index of the nth observation, x_j is the jth column of matrix $X_{N \times K}$, and ω_j is the final weight of the jth variable.

Equation 8 shows the composite index, CI_n, is proportional to the principal components' values.

$$CI_n = \frac{\sum_{i=1}^{K} \lambda_i PC_i}{\sum_{i=1}^{K} \lambda_i}, \qquad CI_n = \frac{\sum_{i=1}^{K} \sum_{j=1}^{K} \lambda_i v_j^i x_j}{\sum_{i=1}^{K} \lambda_i} \tag{8}$$

Hence, it is straightforward to construct the composite index, CI_n based on the PC values [53]. Accordingly, this study examines two 234-length series of CI_n for the trading partners' import and export networks, and two 154-length series of CI_n for the reporting countries' import and export networks.

The goal of this subsection is to compare the core countries in the reporting countries' and trading partners' networks. After constructing the composite index out of the network structural measures, namely the centrality measures, we need to find the countries that are the main contributors of the trade network structure. In an attempt to find the core countries in the trade networks, entropy concept is applied on the composite index. Consequently, we argue that if the core countries in the trading partners' networks are

the same as the core countries in the reporting countries' networks, the structure of these two sets of trade networks would be similar.

Entropy: This concept is an appropriate measure that determines the most variations being captured in the observations [54]. Entropy was first introduced in the field of thermodynamics in the nineteenth century. Later, Shannon (1948) used entropy in the field of information theory [55]. Entropy is the main measure in the information theory, and it estimates how much information is produced in an information source [56].

Shannon (1948) presented a discrete information source as a Markoff process. Based upon this, supposed there is a set of possible events with $p_1, p_2, \ldots, p_n$ probabilities of occurrence. In order to find a measure that captures how much choice/uncertainty/disorder- or chiefly, how much information- is involved in the selection of the event, Shannon proved that there is a measure, say $H(p_1, p_2, \ldots, p_n)$, as a measure of information in information theory. This measure has a form like Eq. (9).

$$H = -M \sum_{i=1}^{n} p_i \log p_i \qquad (9)$$

where M is a positive constant. Accordingly, he called Eq. (9) as the *entropy* of probabilities $p_1, p_2, \ldots, p_n$. Besides, Shannon stated that p_i can be defined as the average frequency of event i, f_i [56].

Note that the information can be stored in variables which can have different values [55]. For the case of Shannon, the information is stored in strings of symbols, but it can be stored in any other kind of variables. Similarly, Skillicorn introduced the entropy concept in the context of dataset, and dimension reduction, particularly, singular value decomposition (SVD). Suppose the dataset properties are presented in a space of r dimensions and we want to represent the dataset in a space of dimension k where $(k \leq r)$. Skillicorn [54] states that one way to find k, is to discuss the contribution of each singular value to the whole, which is computed by Eq. (10).

$$f_k = \frac{s_k^2}{\sum_{i=1}^{r} s_i^2} \qquad (10)$$

where s_i is the ith singular value. By definition, the proposed contribution of each singular value is equivalent to the average frequency of each singular value. Skillicorn [54] states that "entropy measures the amount of disorder in a set of objects", and calculates the entropy of the dataset as Eq. (11) which is equivalent to the entropy measure proposed by Shannon (1948) [54, 56].

$$entropy = \frac{-1}{\log r} \sum_{k=1}^{r} f_k \log(f_k) \qquad (11)$$

In the context of this study, we aimed at finding the main contributors to the trade networks, i.e., the core countries that capture the most information on structural characteristics of the trade networks. Like the messages in Shannon (1948), and the singular values in Skillicorn [54], the composite index of structural network measures associated with countries (*CI*) is considered as the variable storing the information on the structural characteristics of the trade networks. Accordingly, the contribution of the

composite index of structural network measures associated with N countries is calculated as Eq. (12), which is similar to Eq. (10) proposed by Skillicorn [54].

$$f_n = \frac{CI_n^2}{\sum_{i=1}^{N} CI_i^2} \tag{12}$$

where CI_n is the composite index out of the centrality measures for the nth country, $N = 154$ for reporting countries' and $N = 234$ for trading partners' trade (import/export) networks. So, f_n denotes the contribution of the nth country in the considered trade network [54]. The next step is to calculate the entropy of the dataset as Eq. (13), which is similar to Eq. (11) proposed by Skillicorn [54].

$$0 < \text{entropy} = \frac{-1}{\log N} \sum_{n=1}^{N} f_n \log(f_n) < 1 \tag{13}$$

Entropy has a value between 0 (all variation in the dataset is captured in the first country) and 1 (all countries are equally important). The magnitude of the entropy infers how many (and which) counties are the main contributors of the network [54, 57]. In an attempt to find the main contributors, following Zekri et al. [57], first, the *CI* associated with N countries are sorted in descending order. Second, the entropy calculated by Eq. (13) is multiplied by N as follows.

$$\#\quad Contributor \quad Countries = N \times entropy \tag{14}$$

Finally, the main contributor countries of the trade network structure, or equivalently the core countries, are the top number of contributor countries [calculated by Eq. (14)] according to their *CI* being sorted in descending order.

The step-wise network approach to study the robustness of studying trading partners' network adopted in this study is outlined in Fig. 1.
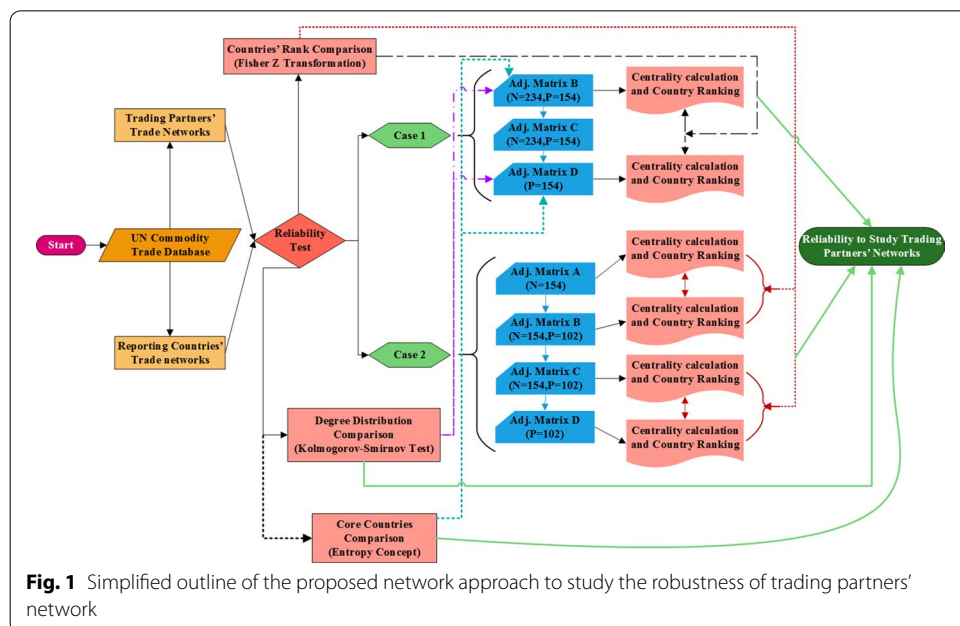

Fig. 1 Simplified outline of the proposed network approach to study the robustness of trading partners' network

Sajedianfard *et al. Appl Netw Sci*   (2021) 6:46

Page 14 of 26

Table 1 Correlation coefficients between the ranks of the 154 countries in the 154-country network and the same countries in the 234-country network, based on centrality measures in 10 commodity groups in the import sector

Correlation coefficients (CC) of Comparing Matrix D in Eq. (5) versus Matrix B in Eq. (5) with $N = 234$ and $P = 154$.

Commodity group	CC (in-degree)	CC (PageRank)	CC (authority)
0	1.00	0.99	1.00
1	1.00	0.93	1.00
2	1.00	0.99	1.00
3	1.00	0.96	1.00
4	1.00	0.94	1.00
5	1.00	0.98	1.00
6	1.00	0.99	1.00
7	1.00	0.98	1.00
8	1.00	0.98	1.00
9	1.00	0.97	1.00

Table 2 Correlation coefficients between the ranks of the 154 countries in the 154-country network and the same countries in the 234-country network, based on centrality measures in 10 commodity groups in the export sector

Correlation coefficients (CC) of Comparing Matrix D in Eq. (5) versus Matrix B in Eq. (5) with $N = 234$ and $P = 154$

Commodity group	CC (out-degree)	CC (PageRank)	CC (hub)
0	1.00	0.99	1.00
1	1.00	0.94	1.00
2	1.00	0.99	1.00
3	1.00	0.97	1.00
4	0.99	0.95	1.00
5	1.00	0.98	1.00
6	1.00	0.99	1.00
7	1.00	0.98	1.00
8	1.00	0.99	1.00
9	0.99	0.99	1.00

Results and discussion

In the following subsection, the effect of missing data on degree and centrality measures is tested. In the second subsection, the core of the reporting countries' and trading partners' networks are compared in import and export sectors using entropy. Lastly, the degree distributions of the reporting countries' and trading partners' networks are compared.

Using centrality measures to test the robustness of studying the trading partners' network

As mentioned in the Database and Methodology section, the examination of the robustness of studying the 234-country network based on 234- and 154-country data is done in two cases. In Case one, the 154-by-154 adjacency matrix and the 234-by-234 adjacency matrix are considered. In Case two, the 154-by-154 adjacency matrix is intentionally rendered to a 154-by-154 adjacency matrix with 102 nonzero columns [matrix B in

Eq. ([5](#)) with $N = 234$ and $P = 102$] which has the same portion of columns with missing data when comparing the 154-country network and 234-country network. Then, this matrix is resized to a full 102-by-102 adjacency matrix which considers solely the trade relation between those 102 countries.

Case 1 In this case, the correlation coefficient of comparing the rank of 154 countries in 154-country network and 234-country network are measured. Countries are ranked based on in-degree, PageRank, and authority in the import networks and based on out-degree, PageRank, and hubness in the export networks. Note that other centrality measures like betweenness, and closeness are not studied in this paper due to the lack of information on the cost of trade between countries. The correlation coefficients for this group of matrices and the ten commodity groups in import and export sectors are shown in Tables [1](#) and [2](#), respectively.

Tables [1](#) and [2](#) represent the correlation coefficients between the ranks of 154 selected countries based on different centrality measures of the 154-country network and the 234-country network in import and export networks, respectively. Highly correlated centrality measures indicate that ranks of the 154 countries common to these two networks do not change significantly, indicating that their situation is negligibly affected by the presence of other countries (i.e., the non-reporting countries). It also shows that zeros in the adjacency matrix do not significantly affect the results of structural measures in the networks under study. These issues are valid for *all* commodity groups in import and export sectors.

Figure [2](#) illustrates the results in Table [1](#). As it is shown, there are less deviation from the blue lines, indicating no significant change in the rank of countries in the 154-country and the 234-country trade networks. This conclusion is valid for *all* the commodity groups. Results in Table [2](#) are illustrated in the "[Appendix](#)".

Case 2 As was stated in the Database and Methodology section, in order to examine the robustness of the results in Case one, a modified adjacency matrix with the same fraction of missing data is constructed; a 154-by-154 adjacency matrix which has 102 nonzero columns. In addition, a full 102-by-102 adjacency matrix based on these nonzero columns and the corresponding rows is constructed. Then, the correlation coefficients between centrality measures of the full 102-country network and the reduced 154-country network, and also the correlation coefficients between centrality measures of the reduced 154-country network and the complete 154-country network for the 102 common countries are calculated. The correlation coefficients of comparing adjacency matrices in this case for the ten commodity groups and the import and export sectors are presented in "[Appendix](#)". Additionally, the similarity between two correlation coefficients in the two pairs of adjacency matrices is tested at the significance level (α) less than 0.01. Accordingly, the *p* value of the Fisher Z transformation for the ten commodity groups and the import and export sectors are shown in "[Appendix](#)".

The same results as the first group of adjacency matrices can be seen for the second group, indicating that this level of missing data does not significantly affect the results in *all* commodity groups. Consequently, it is possible to rely on the results of the 234-country network. In order to check the similarity between the correlation coefficients, the Fisher Z transformation is used to test the the similarity between correlation coefficients

in the two networks. The *p* values from this test are presented in "Appendix" for import and export sectors.

The *p* values for *all* the commodity groups are close to 1 indicating that the probability of similar correlation coefficients is high. Hence it can be concluded that the rank of the countries does not change significantly, so the structure of the trade network in all the commodity groups is not strongly affected by the missing data. This could be due to the scale-free property of trade networks [5, 6, 25, 58–60] which implies the role of the leading or major trading countries in these networks that always report their trade flows.

Scale-free property comes from the fact that small number of nodes have high degree and a large number of them have low degree. Simply put, removing a large number of small countries does not collapse trade networks. For this purpose, the case 2 is carried out once more in which the number of countries taken into account is now 68 rather than 102. The correlation coefficients of comparing adjacency matrices in this case for 10 commodity groups and the import and export sectors are presented in "Appendix".

As before, it can be concluded that the structure of trade networks in *all* the commodity groups is formed by the leading and major participating countries in trade which always report their trade-flow statistics. Accordingly, the structure of the network is not significantly affected by the missing data.

Core countries in the reporting countries' and trading partners' networks

Centrality measures capture the local and global structural features of countries in trade networks. In an attempt to study the structure of the reporting countries' and trading partners' networks, a composite index out of centrality measures is constructed using PCA. For the import sector, the in-degree, PageRank, and authority centralities and for the export sector, the out-degree, PageRank, and hubness measures are considered. Lastly, the core countries of the ten commodity import and export networks are extracted using entropy. The advantage of this approach is that we do not intentionally reduce the size of the matrices to analyze the network structures and we rely on the information provided by the dataset.

The results show that the core of the reporting countries' network in trading (import/ export) the ten commodity groups are the totally included in the core of the trading partners' networks. Because the core of the networks are extracted using structural features of countries, the results indicate that the structure of the reporting countries' and trading partners' networks are the same. This confirms the results in previous section.

The common core countries in importing all the ten commodity groups are Belgium, Canada, China, Germany, India, Italy, Japan, Mexico, Taiwan, Poland, Turkey, USA, and UK. Note that Taiwan, the province of China, is reported as Other Asia, not elsewhere specified. Also, the common core countries in exporting the ten commodity groups are Australia, Austria, Belgium, Canada, China, Denmark, France, Germany, India, Italy, Japan, Mexico, Netherlands, Norway, Poland, Republic of Korea, Russian Federation, Saudi Arabia, Singapore, Spain, Sweden, Switzerland, Turkey, USA, UAE, and UK. In the interest of space, the core of the reporting countries' and trading partners' networks in all commodity groups and import/export sectors are not included in the paper and we keep available on request.

Table 3 *p* value of the Kolmogorov–Smirnov test of the upper tail of the strength distribution for 10 commodity groups in the import and export sectors

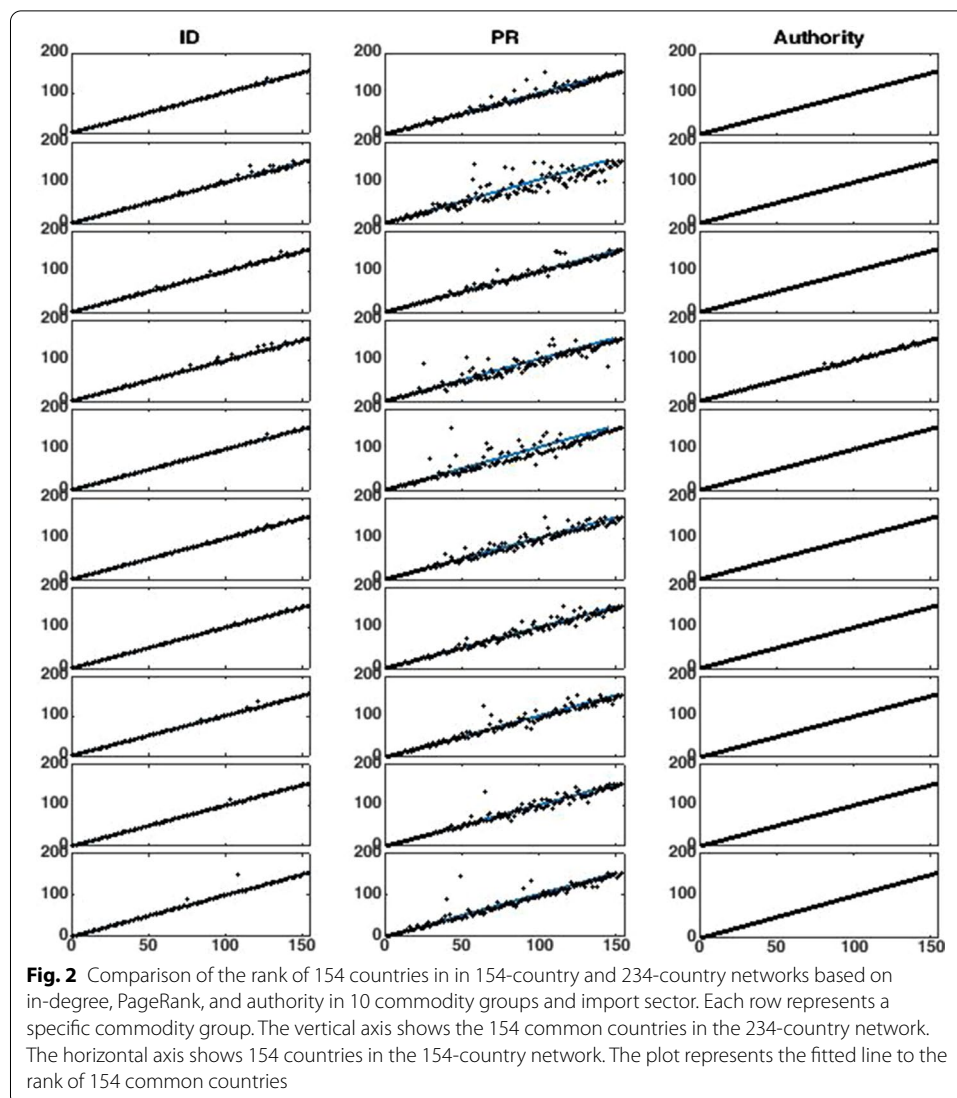| Commodity group	*p* value of the Kolmogorov–Smirnov test		Upper tail of the strength distribution	
	Import sector	Export sector	Import sector	Export sector
0	0.99	0.93	0.2×10^{11}	0.6×10^{11}
1	1.00	1.00	0.2×10^{10}	0.5×10^{10}
2	1.00	0.99	0.2×10^{11}	2×10^{10}
3	0.99	1.00	0.2×10^{11}	2×10^{10}
4	1.00	0.95	1.5×10^{9}	0.2×10^{10}
5	1.00	1.00	0.2×10^{11}	0.5×10^{11}
6	1.00	1.00	0.2×10^{11}	0.5×10^{11}
7	1.00	0.99	1×10^{11}	2×10^{11}
8	0.95	0.99	0.5×10^{11}	0.5×10^{11}
9	0.99	0.99	1×10^{10}	1×10^{10}

The first column lists the commodity groups. The second and third columns describe the *p* value of the Kolmogorov–Smirnov test in the import and export sectors, respectively. The fourth and fifth columns show the upper tail of the strength distribution in the import and export sectors, respectively

Degree distribution of the reporting countries' and trading partners' networks

Another structural aspect of networks can be seen in the degree distribution. Two main types of networks are of interest in this context [4–6, 18, 61]: small-world, and scale-free. Small-world networks are described by clustering coefficient and the characteristic path length properties. Networks with higher clustering and almost the same short average path as random networks are called small-world networks. Scale-free networks are those that have few highly connected nodes and many slightly connected nodes, with a power-law degree distribution [62].

In order to check the similarity of the degree distributions in the 154-country network and the 234-country network, a Kolmogorov–Smirnov test is done [63]. The previous studies of trade networks have argued that these networks are scale-free [5, 6, 25, 58–60]; a large number of nodes with low degree and a small number of nodes with high degree. In other words, the degree distribution in the trade network has a fat tail, indicating that only a small number of countries are hubs in this network; this is a consequence of the scale-free property in network theory [64, 65].
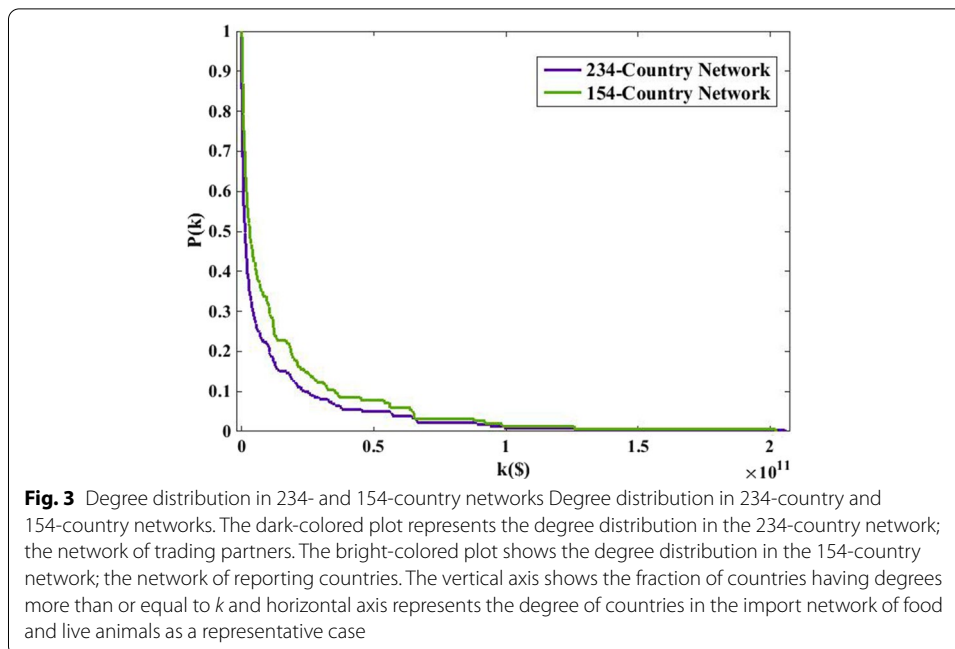
In scale-free networks it is possible to study the tail, i.e., large values of degree. Hence, in this study, after calculating the weighted degrees by using the Brain Connectivity Toolbox in MATLAB [66], the Kolmogorov–Smirnov test was done for the upper tail of the degree distributions which are mentioned in Table 3. The *p* values of the Kolmogorov–Smirnov test for *all* the commodity groups in import and export sectors are presented in Table 3. The results show that the null hypothesis (similar distribution) is not rejected and likely to be true since the *p* value is at least 0.93. These results also show that trade networks in *all* the commodity groups are formed based on the bilateral trade between a tiny number of central countries, indicating the scale-free property of these networks. To put it in another way, although the dominant countries are not always the same depending on the commodity groups, the top countries are present as top importers and/or exporters in most/all groups.

Fig. 2 Comparison of the rank of 154 countries in in 154-country and 234-country networks based on in-degree, PageRank, and authority in 10 commodity groups and import sector. Each row represents a specific commodity group. The vertical axis shows the 154 common countries in the 234-country network. The horizontal axis shows 154 countries in the 154-country network. The plot represents the fitted line to the rank of 154 common countries

Besides, Fig. 3 shows the degree distribution in both trading partners' and reporting countries' import networks of food and live animals as a representative example of all the trade networks examined in this study. This figure illustrates the overall similarity between the degree distributions in the reporting countries' and the trading partners' networks.

Conclusions

The complexity of trading relations and their inter-country distributions within the world trade network has been successfully captured and analyzed in literature by applying network concepts. Previous studies mainly followed the common approach which takes into account either the bilateral trade between a specific group of countries or between countries that actually report trade flows. That is to say, those trade networks had lack of information about the trade between all the countries in the whole world. This is while, it is important to study an expanded trade data that contains a more complete information set about bilateral trades rather than the available

Fig. 3 Degree distribution in 234- and 154-country networks Degree distribution in 234-country and 154-country networks. The dark-colored plot represents the degree distribution in the 234-country network; the network of trading partners. The bright-colored plot shows the degree distribution in the 154-country network; the network of reporting countries. The vertical axis shows the fraction of countries having degrees more than or equal to k and horizontal axis represents the degree of countries in the import network of food and live animals as a representative case

data. This expanded data is the basis of studying a network close to the *real* international trade. In order to fill the gap in previous studies, a trade network based on all the countries and territories in the world trade is constructed and the robustness of studying this imputed network is examined in this article.

Exploration of what constitute suitable data, how to deal with missing data, and demonstrating the initial results using key network measures are investigated in this study. Specifically, it has presented a systematic way to examine the robustness of studying a network based on trading partners of reporting countries, which has more information about bilateral trade in the world trade than previous studies. In particular, measures such as centrality and degree distributions are used to verify the robustness of studying the trading partners' network. The core of this study is to compare the structure of networks with missing data and without missing data. The results are illustrated and verified using the import and export of the ten commodity groups. It must be noted that although a country may have different rank in different commodity groups, this study *does not* compare the rank of a specific country among different commodity groups.
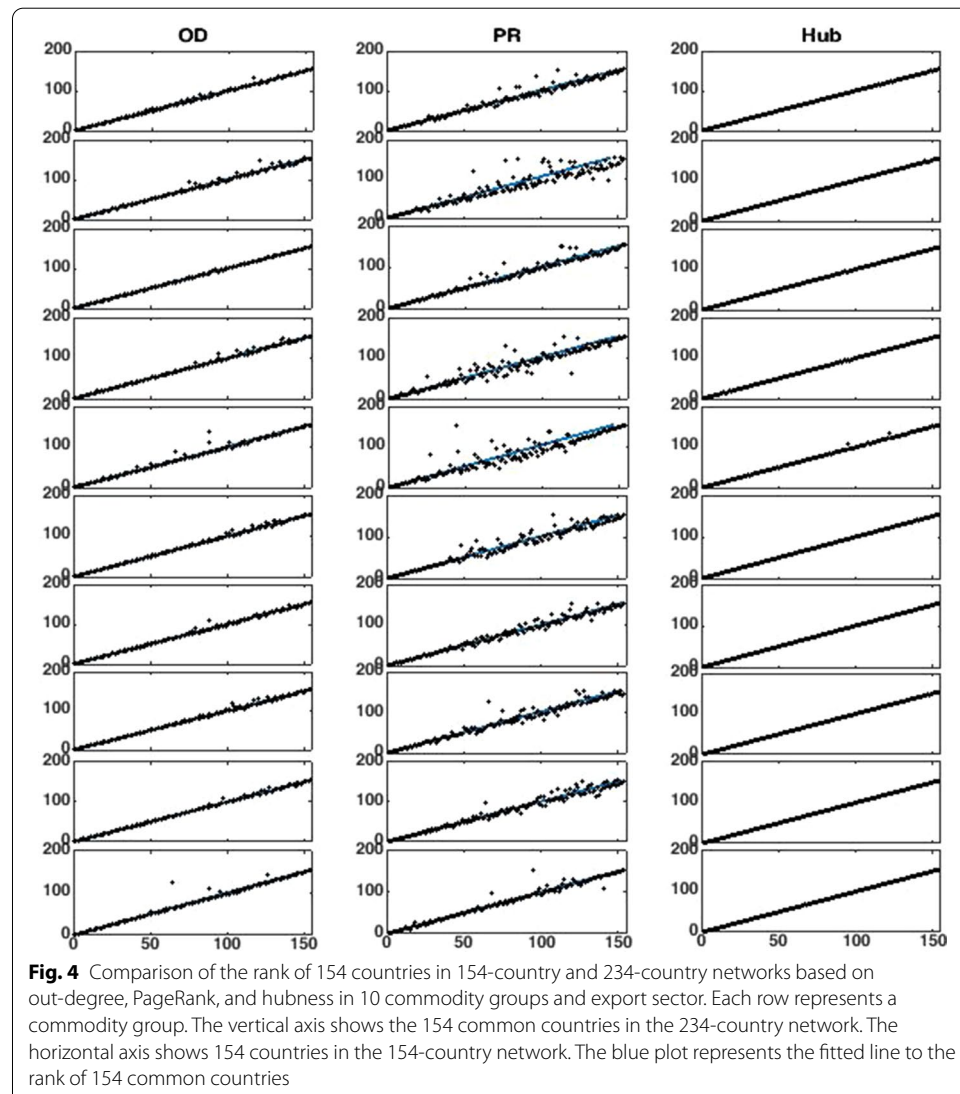
The results of this study show that the trading partners' network is robust to study in *all* the commodity groups and import and export sectors. That is to say, we found that leaving approximately half of the countries out from the reporting countries' network (which has lack of information on bilateral trade in world) did not make a significant difference relative to the trade network based on trading partners of reporting countries. So, we infer that in this case missing data does not affect the overall structure of the network and verify that the leading countries determine the overall network structure. These findings are valid for all the commodity groups and import and export sectors. Accordingly, it is possible to study a network containing more information that is important in the view point of all types of countries for finding potential trading partners or other policy purposes.

This research examined a fundamental topic in studying trade networks which contributes a basis to future studies. The robustness of studying an expanded trade network provides a wider insight into the first stages of countries' foreign trade policy making. Particularly, the advantage of studying the expanded trade network is that all types of countries, regardless of their size, can evaluate their own, and their partners' positions in trade networks. A more accurate examination of current trade status can be helpful to policy makers in each country to modify their international trade policy and strategies.

Appendix

Comparison of the rank of 154 countries in 154-country and 234-country networks in export sector

Figure 4 illustrates the correlation plots corresponding to the comparison of the rank of 154 countries in 154-country and 234-country networks in export sector.

Fig. 4 Comparison of the rank of 154 countries in 154-country and 234-country networks based on out-degree, PageRank, and hubness in 10 commodity groups and export sector. Each row represents a commodity group. The vertical axis shows the 154 common countries in the 234-country network. The horizontal axis shows 154 countries in the 154-country network. The blue plot represents the fitted line to the rank of 154 common countries

Correlation coefficient comparison of 102 countries

The correlation coefficients between the rank of 102 countries in various commodity groups and the import and export sectors are presented in Tables 4 and 5, respectively.

Fisher Z transformation

The *p* value of the Fisher Z Transformation for 10 commodity groups and the import and export sectors are shown in Tables 6 and 7, respectively.

Correlation coefficient comparison of 68 countries

The correlation coefficients between the rank of 68 countries in various commodity groups and the import and export sectors are presented in Tables 8 and 9, respectively.

Table 4 Correlation coefficients between the rank of 154 countries in full 154-country network and reduced 154-country network [part (a)], and correlation coefficients between the rank of 102 countries in reduced 154-country network and 102-country network [part (b)] based on centrality measures in import sector

Commodity group	CC (in-degree)	CC (PageRank)	CC (authority)
(a): Correlation coefficients (CC) of Comparing Matrix D in Eq. (5) with $P = 102$ versus Matrix B in Eq. (5) with $N = 154$ and $P = 102$			
0	1.00	1.00	1.00
1	1.00	0.98	1.00
2	1.00	0.99	1.00
3	0.99	0.98	1.00
4	0.99	0.99	1.00
5	1.00	0.99	1.00
6	1.00	0.99	1.00
7	1.00	0.99	1.00
8	1.00	0.99	1.00
9	1.00	0.99	1.00
(b): Correlation coefficients (CC) of Comparing Matrix B in Eq. (5) with $N = 154$ and $P = 102$ versus Matrix A in Eq. (5) with $N = 154$			
0	1.00	1.00	1.00
1	1.00	0.99	1.00
2	1.00	1.00	1.00
3	0.99	0.98	1.00
4	0.99	1.00	1.00
5	1.00	1.00	1.00
6	1.00	1.00	1.00
7	1.00	1.00	1.00
8	1.00	1.00	1.00
9	1.00	1.00	1.00

The first column lists the 10 commodity groups. The second, third, and fourth columns describe the correlation coefficients for different pairs of networks mentioned in part (a) and (b), based on in-degree, PageRank, and authority, respectively

Sajedianfard *et al. Appl Netw Sci*   (2021) 6:46

Page 22 of 26

Table 5 Correlation coefficients between the rank of 154 countries in full 154-country network and reduced 154-country network [part (a)], and correlation coefficients between the rank of 102 countries in reduced 154-country network and the 102-country network [part (b)] based on centrality measures in export sector

Commodity group	CC (out-degree)	CC (PageRank)	CC (hub)
(a): Correlation coefficients (CC) of Comparing Matrix D in Eq. (5) with $P = 102$ versus Matrix B in Eq. (5) with $N = 154$ and $P = 102$			
0	0.99	0.99	1.00
1	0.98	0.92	1.00
2	1.00	1.00	1.00
3	1.00	1.00	1.00
4	0.98	0.99	1.00
5	1.00	0.99	1.00
6	1.00	1.00	1.00
7	0.99	1.00	1.00
8	0.99	0.99	1.00
9	1.00	1.00	1.00
(b): Correlation coefficients (CC) of Comparing Matrix B in Eq. (5) with $N = 154$ and $P = 102$ versus Matrix A in Eq. (5) with $N = 154$			
0	1.00	0.97	1.00
1	1.00	0.95	1.00
2	1.00	0.98	1.00
3	1.00	0.90	1.00
4	1.00	0.94	1.00
5	1.00	0.97	1.00
6	1.00	0.95	1.00
7	1.00	0.95	1.00
8	1.00	0.97	1.00
9	1.00	0.99	1.00

The first column lists the 10 commodity groups. The second, third, and fourth columns describe the correlation coefficients for different pairs of networks mentioned in part (a) and (b), based on out-degree, PageRank, and hub, respectively

Table 6 Test of the similarity between two correlation coefficients in the two pairs of networks in case 2, in import sector

p value of the similarity between correlation coefficients at $\alpha < 0.01$ in import sector			
Commodity group	In-degree	PageRank	Authority
0	1.00	1.00	1.00
1	1.00	1.00	1.00
2	1.00	1.00	1.00
3	1.00	1.00	1.00
4	1.00	1.00	1.00
5	1.00	1.00	1.00
6	1.00	1.00	1.00
7	1.00	1.00	1.00
8	1.00	1.00	1.00
9	1.00	1.00	1.00

The first column lists the 10 commodity groups. The second, third, and fourth columns describe the *p* value of Fisher Z Transformation based on centrality measures. These network measures are in-degree, PageRank, and authority, respectively

Sajedianfard *et al. Appl Netw Sci*   (2021) 6:46

Page 23 of 26

Table 7 Test of the similarity between two correlation coefficients in the two pairs of networks in case 2, in export sector

Commodity group	Out-degree	PageRank	Hub
p value of the similarity between correlation coefficients at *α* < 0.01 in export sector			
0	1.00	1.00	1.00
1	1.00	1.00	1.00
2	1.00	1.00	1.00
3	1.00	1.00	1.00
4	1.00	1.00	1.00
5	1.00	1.00	1.00
6	1.00	1.00	1.00
7	1.00	1.00	1.00
8	1.00	1.00	1.00
9	1.00	1.00	1.00

The first column lists the 10 commodity groups. The second, third, and fourth columns describe the *p* value of Fisher Z Transformation based on centrality measures. These network measures are out-degree, PageRank, and hub, respectively

Table 8 Correlation coefficients between the rank of 102 countries in full 102-country network and reduced 102-country network [part (a)], and correlation coefficients between the rank of 68 countries in reduced 102-country network and the 68-country network [part (b)], based on centrality measures in import sector

Commodity group	CC (in-degree)	CC (PageRank)	CC (authority)
(a): Correlation coefficients (CC) of Comparing Matrix *D* in Eq. (5) with *P* = 68 versus Matrix *B* in Eq. (5) with *N* = 102 and *P* = 68			
0	0.99	1.00	1.00
1	1.00	0.99	1.00
2	1.00	1.00	1.00
3	0.98	0.99	0.99
4	0.97	0.99	0.98
5	1.00	0.99	1.00
6	1.00	1.00	1.00
7	1.00	1.00	1.00
8	1.00	1.00	1.00
9	0.99	1.00	1.00
(b): Correlation coefficients (CC) of Comparing Matrix Matrix *B* in Eq. (5) with *N* = 102 and *P* = 68 versus Matrix *A* in Eq. (5) with *N* = 102			
0	1.00	1.00	1.00
1	1.00	0.98	1.00
2	1.00	1.00	1.00
3	1.00	0.99	1.00
4	1.00	0.99	1.00
5	1.00	1.00	1.00
6	1.00	1.00	1.00
7	1.00	1.00	1.00
8	1.00	1.00	1.00
9	1.00	1.00	1.00

The first column lists the 10 commodity groups. The second, third, and fourth columns describe the correlation coefficients for the different pairs of networks mentioned in part (a) and (b), based on in-degree, PageRank, and authority, respectively

Table 9 Correlation coefficients between the rank of 102 countries in full 102-country network and reduced 102-country network [part (a)], and correlation coefficients between the rank of 68 countries in reduced 102-country network and the 68-country network [part (b)], based on centrality measures in export sector

Commodity group	CC (out-degree)	CC (PageRank)	CC (hub)
(a): Correlation coefficients (CC) of Comparing Matrix D in Eq. (5) with $P = 68$ versus Matrix B in Eq. (5) with $N = 102$ and $P = 68$			
0	1.00	1.00	1.00
1	0.98	0.99	1.00
2	1.00	1.00	1.00
3	0.99	0.99	1.00
4	0.99	0.99	1.00
5	1.00	1.00	1.00
6	1.00	1.00	1.00
7	1.00	1.00	1.00
8	0.99	1.00	1.00
9	0.99	0.99	1.00
(b): Correlation coefficients (CC) of Comparing Matrix Matrix B in Eq. (5) with $N = 102$ and $P = 68$ versus Matrix A in Eq. (5) with $N = 102$			
0	1.00	0.99	1.00
1	1.00	0.82	1.00
2	1.00	1.00	1.00
3	1.00	0.95	1.00
4	1.00	0.96	1.00
5	1.00	0.96	1.00
6	1.00	0.96	1.00
7	1.00	0.99	1.00
8	1.00	0.96	1.00
9	1.00	0.97	1.00

The first column lists the 10 commodity groups. The second, third, and fourth columns describe the correlation coefficients for the different pairs of networks mentioned in part (a) and (b), based on out-degree, PageRank, and hub, respectively

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Economics, Shiraz University, Shiraz, Iran. [2]School of Architecture, Design and Planning, The University of Sydney, New South Wales 2006, Australia. [3]School of Physics, The University of Sydney, New South Wales 2006, Australia. [4]Center for Integrative Brain Function, The University of Sydney, New South Wales 2006, Australia.

References
1. Wasserman S, Faust K (1994) Social network analysis: methods and applications, vol 8. Cambridge University Press, Cambridge
2. Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45(2):167–256. https://doi.org/10.1137/S003614450342480
3. Jackson MO (2010) Social and economic networks. Princeton University Press, Princeton
4. Serrano MA, Boguná M (2003) Topology of the world trade web. Phys Rev E 68(1):015101. https://doi.org/10.1103/PhysRevE.68.015101
5. Li X, Jin YY, Chen G (2003) Complexity and synchronization of the world trade web. Physica A 328(1–2):287–296. https://doi.org/10.1016/S0378-4371(03)00567-3
6. Garlaschelli D, Loffredo MI (2004) Fitness-dependent topological properties of the world trade web. Phys Rev Lett 93(18):188701. https://doi.org/10.1103/PhysRevLett.93.188701
7. Baskaran T, Brück T (2005) Scale-free networks in international trade. Technical report, DIW Discussion Papers. http://www.diw.de/documents/publikationen/73/diw_01.c.43299.de/dp493.pdf
8. Fagiolo G, Reyes J, Schiavo S (2008) On the topological properties of the world trade web: a weighted network analysis. Physica A 387(15):3868–3873. https://doi.org/10.1016/j.physa.2008.01.050
9. Fagiolo G (2010) The international-trade network: gravity equations and topological properties. J Econ Interact Coord 5(1):1–25. https://doi.org/10.1007/s11403-010-0061-y
10. Squartini T, Fagiolo G, Garlaschelli D (2011) Randomizing world trade. I. A binary network analysis. Phys Rev E 84(4):046117. https://doi.org/10.1103/PhysRevE.84.046117
11. Squartini T, Fagiolo G, Garlaschelli D (2011) Randomizing world trade. II. A weighted network analysis. Phys Rev E 84(4):046118. https://doi.org/10.1103/PhysRevE.84.046118
12. De Benedictis L, Nenci S, Santoni G, Tajoli L, Vicarelli C (2014) Network analysis of world trade using the BACI-CEPII dataset. Glob Econ J 14(3–4):287–343. https://doi.org/10.1515/gej-2014-0032
13. Deguchi T, Takahashi K, Takayasu H, Takayasu M (2014) Hubs and authorities in the world trade network using a weighted hits algorithm. PLoS ONE. https://doi.org/10.1371/journal.pone.0100338.g001
14. Abbate A, De Benedictis L, Fagiolo G, Tajoli L (2018) Distance-varying assortativity and clustering of the international trade network-ADDENDUM. Netw Sci 6(4):633–633. https://doi.org/10.1017/nws.2018.16
15. de Andrade RL, Rêgo LC (2018) The use of nodes attributes in social network analysis with an application to an international trade network. Physica A 491:249–270. https://doi.org/10.1016/j.physa.2017.08.126
16. Ding H, Jin Y, Liu Z, Xie W (2019) The relationship between international trade and capital flow: a network perspective. J Int Money Finance 91:1–11. https://doi.org/10.1016/j.jimonfin.2018.10.001
17. Yan B, Luo J (2019) Multicores-periphery structure in networks. Netw Sci 7(1):70–87. https://doi.org/10.1017/nws.2018.27
18. Garlaschelli D, Loffredo MI (2005) Structure and evolution of the world trade network. Physica A 355(1):138–144. https://doi.org/10.1016/j.physa.2005.02.075
19. Bhattacharya K, Mukherjee G, Saramäki J, Kaski K, Manna SS (2008) The international trade network: weighted network analysis and modelling. J Stat Mech Theory Exp 2008(02):02002. https://doi.org/10.1088/1742-5468/2008/02/P02002
20. Reyes J, Schiavo S, Fagiolo G (2008) Assessing the evolution of international economic integration using random walk betweenness centrality: the cases of east asia and latin america. Adv Complex Syst 11(05):685–702. https://doi.org/10.1142/S0219525908001945
21. Tzekina I, Danthi K, Rockmore DN (2008) Evolution of community structure in the world trade web. Eur Phys J B 63(4):541–545. https://doi.org/10.1140/epjb/e2008-00181-2
22. Zhang J, Cui Z, Zu L (2014) The evolution of free trade networks. J Econ Dyn Control 38:72–86. https://doi.org/10.1016/j.jedc.2013.09.004
23. Zhu Z, Cerina F, Chessa A, Caldarelli G, Riccaboni M (2014) The rise of china in the international trade network: a community core detection approach. PLoS ONE 9(8):105496. https://doi.org/10.1371/journal.pone.0105496
24. Matous P, Todo Y (2016) Energy and resilience: the effects of endogenous interdependencies on trade network formation across space among major japanese firms. Netw Sci 4(2):141–163. https://doi.org/10.1017/nws.2015.37
25. Zhou M, Wu G, Xu H (2016) Structure and formation of top networks in international trade, 2001–2010. Soc Netw 44:9–21. https://doi.org/10.1016/j.socnet.2015.07.006
26. del Río-Chanona RM, Grujić J, Jensen HJ (2017) Trends of the world input and output network of global trade. PLoS ONE 12(1):0170817. https://doi.org/10.1371/journal.pone.0170817
27. Fracasso A, Nguyen HT, Schiavo S (2018) The evolution of oil trade: a complex network approach. Netw Sci 6(4):545–570. https://doi.org/10.1017/nws.2018.6
28. Gleditsch KS (2002) Expanded trade and GDP data. J Conflict Resolut 46(5):712–724. https://doi.org/10.1177/0022002702046005006

29. The World Bank group (2021) Total population. https://data.worldbank.org/indicator/SP.POP.TOTL
30. Huang S, Gou W, Cai H, Li X, Chen Q (2020) Effects of regional trade agreement to local and global trade purity relationships. Complexity. https://doi.org/10.1155/2020/2987217
31. Newman MEJ (2010) Networks: an introduction. Oxford University Press, Oxford
32. Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. Rev Mod Phys 74(1):47. https://doi.org/10.1103/RevModPhys.74.47
33. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D-U (2006) Complex networks: structure and dynamics. Phys Rep 424(4–5):175–308. https://doi.org/10.1016/j.physrep.2005.10.009
34. Restrepo JG, Ott E, Hunt BR (2007) Approximating the largest eigenvalue of network adjacency matrices. Phys Rev E 76(5):056119. https://doi.org/10.1103/PhysRevE.76.056119
35. Watts DJ, Strogatz SH (1998) Collective dynamics of "small-world" networks. Nature 393(6684):440. https://doi.org/10.1038/30918
36. Atay FM, Biyikoglu T, Jost J (2006) Synchronization of networks with prescribed degree distributions. IEEE Trans Circuits Syst I Regul Pap 53(1):92–98. https://doi.org/10.1109/TCSI.2005.854604
37. Newman ME, Watts DJ (1999) Renormalization group analysis of the small-world network model. Phys Lett A 263(4–6):341–346. https://doi.org/10.1016/S0375-9601(99)00757-4
38. Barabási A-L, Bonabeau E (2003) Scale-free networks. Sci Am 288(5):60–69. https://doi.org/10.1038/scientificamerican0503-60
39. Jackson MO, Rogers BW, Zenou Y (2017) The economic consequences of social-network structure. J Econ Lit 55(1):49–95. https://doi.org/10.1257/jel.20150694
40. Franceschet M (2014) Katz centrality. https://www.sci.unich.it/~francesc/teaching/network/katz.html
41. Dijkstra EW et al (1959) A note on two problems in connexion with graphs. Numer Math 1(1):269–271. https://doi.org/10.1007/BF01386390
42. Newman ME (2001) Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Phys Rev E 64(1):016132. https://doi.org/10.1103/PHYSREVE.64.016132
43. United Nations (2019) United Nations Comtrade database. https://comtrade.un.org/data/
44. United Nations (2016) Every user of United Nations Comtrade should know the coverage and limitations of the data. https://comtrade.un.org/db/help/ureadMeFirst.aspx
45. United Nations (2011) International merchandise trade statistics (IMTS), concepts and definitions (2010). https://unstats.un.org/unsd/trade/eg-imts/IMTS%202010%20(English).pdf
46. Kendall MG (1970) Rank correlation methods. Griffin, London
47. Chalmer BJ (2020) Understanding statistics. CRC Press, Boca Raton
48. Salkind NJ (2007) Fisher's Z transformation. Encyclopedia of measurement and statistics, vol 1. SAGE Publications, Inc., Thousand Oaks, pp 361–364. https://doi.org/10.4135/9781412952644.n175
49. Jolliffe IT (2002) Principal component analysis. Springer, Gateway East
50. Abdi H, Williams LJ (2010) Principal component analysis. Wiley Interdiscip Rev Comput Stat 2(4):433–459. https://doi.org/10.1002/wics.101
51. Sunderland KM, Beaton D, Fraser J, Kwan D, McLaughlin PM, Montero-Odasso M, Peltsch AJ, Pieruccini-Faria F, Sahlas DJ, Swartz RH et al (2019) The utility of multivariate outlier detection techniques for data quality evaluation in large studies: an application within the ondri project. BMC Med Res Methodol 19(1):1–16. https://doi.org/10.1186/s12874-019-0737-5
52. Vyas S, Kumaranayake L (2006) Constructing socio-economic status indices: how to use principal components analysis. Health Policy Plan 21(6):459–468. https://doi.org/10.1093/heapol/czl029
53. Chen B, Woo YP (2010) Measuring economic integration in the Asia-Pacific region: a principal components approach. Asian Econ Pap 9(2):121–143. https://doi.org/10.1162/ASEP_a_00009
54. Skillicorn D (2019) Understanding complex datasets: data mining with matrix decompositions. CRC Press, Boca Raton
55. Sethneha (2020) Entropy—a key concept for all data science beginners. https://www.analyticsvidhya.com/blog/2020/11/entropy-a-key-concept-for-all-data-science-beginners
56. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27(3):379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
57. Zekri H, Mokhtari AR, Cohen DR (2016) Application of singular value decomposition (SVD) and semi-discrete decomposition (SDD) techniques in clustering of geochemical data: an environmental study in central Iran. Stoch Environ Res Risk Assess 30(7):1947–1960. https://doi.org/10.1007/s00477-016-1219-5
58. Kali R, Reyes J (2007) The architecture of globalization: a network approach to international economic integration. J Int Bus Stud 38(4):595–620. https://doi.org/10.1057/palgrave.jibs.8400286
59. Baskaran T, Blöchl F, Brück T, Theis FJ (2011) The Heckscher–Ohlin model and the network structure of international trade. Int Rev Econ Finance 20(2):135–145. https://doi.org/10.1016/j.iref.2010.11.003
60. Fagiolo G, Reyes J, Schiavo S (2009) World-trade web: Topological properties, dynamics, and evolution. Phys Rev E 79(3):036115. https://doi.org/10.1103/PhysRevE.79.036115
61. Kaluza P, Kölzsch A, Gastner MT, Blasius B (2010) The complex network of global cargo ship movements. J R Soc Interface 7(48):1093–1103. https://doi.org/10.1098/rsif.2009.0495
62. Fornito A, Zalesky A, Bullmore E (2016) Fundamentals of brain network analysis. Elsevier, Amsterdam
63. Dodge Y (2008) The concise encyclopedia of statistics. Springer, Berlin
64. Barabási A-L (2016) Network science. Cambridge University Press, Cambridge
65. Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási A-L (2007) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci USA 104(18):7332–7336. https://doi.org/10.1073/pnas.0610245104
66. Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: uses and interpretations. NeuroImage 52(3):1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

Publisher's Note